

Magdalena Edut

Université de Silésie
Katowice

Phénomènes naturels – une esquisse orientée-objets

Abstract

In her article Magdalena Edut presented a new type of the lexicographical description which is based on the methodology of the object-oriented approach (*approche orientée-objets*). In the first part of the text she discussed, on the example of the word *mżawka* (*drizzle*), which belongs to the category of the "natural phenomena", the principles and aims of the proposed approach. In the following parts she compared the received description with the description suggested by the lexical database WordNet and with the contemporary lexicographical theories dealing with the issue of the automatic translation:

- prof. G. Gross's idea of an electronic dictionary,
- I. Mel'čuk's and A. Zolkovsky's model Sens-Texte,
- Pustejovsky's and Boguraev's structure *qualia*.

Keywords

Object-oriented approach, object classes, the *qualia* structure, lexical functions.

Dans cette communication nous proposons la présentation d'une conception nouvelle connue sous le nom d'approche orientée-objets qui s'inscrit dans le cadre du traitement automatique du langage. Dans la première partie nous exposerons les principes et l'objectif de cette approche en s'appuyant sur les exemples appartenant à la classe des *phénomènes naturels* dont les recherches se sont presque toujours limitées à l'analyse des phrases verbales de type : *il pleut, il neige*. Ensuite nous essayerons de voir les rapports et les ressemblances de cette conception avec d'autres théories lexicographiques.

La notion de sens est devenue de plus en plus importante dans le traitement automatique de la langue et les démarches qui séparaient la syntaxe et la sémantique se sont montrées inadéquates et insuffisantes pour une bonne description. La méthode proposée essaiera de réunir des informations séman-

tiques, lexicales, syntaxiques et morphologiques. Elle tentera de donner des informations explicites, qui sont traditionnellement contenues dans les dictionnaires et de tirer les informations implicites. L'objectif de la conception est plutôt de créer une base de données lexicales polyvalente que de construire un dictionnaire sous forme électronique.

La méthode que l'on propose est de type orientée objets, ce qui exige comme point de départ pour l'analyse le choix des objets. Cette façon de procéder distingue d'une manière radicale cette conception de la description en terme prédicat-argument où on part de fonction (prédicat) en vue de trouver ses arguments.

Dans la conception orientée-objets l'objet va être défini par deux types d'informations :

- d'un côté par des objets qui l'accompagnent dans des situations évoquées, alors par les attributs que nous sommes capables d'énumérer à propos de cet objet :

bruine – pénétrante, verglaçante, interminable

vent – violent, fort, desséchant, piquant, pluvieux, froid, humide, vif

pluie – grosse, fine, pénétrante, cinglante, diluvienne, épouvantable

- de l'autre côté par les opérations que l'objet lui-même peut réaliser et celles qui peuvent être effectuées sur cet objet.

Pour systématiser davantage les informations qui caractérisent une entrée, on propose la division des opérations en trois sous-classes :

a) celles qui forment la classe d'objet ou provoquent que la classe n'apparaît pas, appelées *constructeurs* :

(f) *bruine – commencer à tomber, il y a*

(m) *vent – se lever, commencer à souffler*

(f) *pluie – commencer à tomber, apparaître*

b) celles que la classe donnée peut effectuer elle-même, appelées *accesseurs* :

(m) *bruine – tomber, se transformer en pluie*

(m) *vent – souffler, siffler, disperser les nuages*

(f) *pluie – tomber, tomber à seaux, cingler*

c) celles qui donnent les informations sur le comportement, la structure de la classe et qui peuvent être effectuées sur l'objet, appelées *manipulateurs* :

(f) *bruine – prévoir, annoncer*

(m) *vent – être tanné par, être hâlé par*

(f) *pluie – s'abriter de la, se protéger de la*

Puisque l'objectif auquel on vise est de présenter une vision du monde à travers la langue, les caractéristiques ontologiques de l'objet en question ne vont pas être prises en compte. Nous allons nous concentrer sur le comportement linguistique de l'objet, c'est-à-dire présenter son entourage lexical. Sur ce point, la méthode discutée s'oppose à la conception classique de descriptions consistant à une énumération des sens lexicaux – un *listing* exhaustif de toutes les opérations qui s'appliquent à l'objet en question.

La caractérisation a un caractère ouvert, ainsi est-il possible de modifier, d'enrichir la description de l'objet traité. Grâce à l'application du mécanisme d'héritage sémantique les opérations pourront être héritées de la superclasse. Chaque mot, étant membre d'une hiérarchie, possède un nombre de prédictats qui lui sont attribués selon la position qu'il occupe dans cette hiérarchie. Ainsi le mot *bruine* étant une sorte de précipitation va hériter de différentes opérations propres à cette classe telles que: *tomber*, *commencer à tomber*, *cesser*, *apparaître*.

L'intérêt majeur de l'héritage sémantique est de ne pas répéter les attributs et les opérations pour une nouvelle sous-classe, mais les ajouter à ceux déjà existants.

Une telle description nous amènera à présenter le monde tel qu'il est reflété dans la langue. Voyons l'exemple d'une entrée :

PL	FR
mżawka	bruine
[Klasa obiektywowa :] zjawisko naturalne	[Classe d'objets:] phénomène naturel
[Definicja :] drobny, zimny deszcz wynikający z opadu mgły	[Définition:] petite pluie très fine et froide, qui résulte de la précipitation du brouillard
[Synonimy :] mżawka	[Synonymes]: crachin
[Superklasa 5:] zjawisko naturalne	[Superclasse 5:] phénomène naturel
[Superklasa 4:] zjawisko atmosferyczne	Superclasse 4:] phénomène atmosphérique
[Superklasa 3:] zjawisko meteorologiczne	[Superclasse 3:] phénomène météorologique
[Superklasa 2:] opad	[superclasse 2:] précipitation
[Superklasa 1:] deszcz	[Superclasse 1:] pluie
[Domena:]	[Domaine:]
[Atrybuty:]	[Attributs:]
<i>marznąca mżawka</i>	<i>la bruine verglaçante</i>
<i>nieustająca mżawka</i>	<i>une bruine interminable</i>
<i>poranna mżawka</i>	<i>une bruine matinale</i>
<i>przenikliwa mżawka</i>	<i>une bruine pénétrante</i>
[Część-całość:]	[Partie-tout:]
[Operacje:]	[Opérations:]

PL	FR
mżawka	bruine
[Operacje konstruktor :]	[Opérations constructeur :]
<i>mżawka zaczyna padać</i>	<i>la bruine commence à tomber</i>
<i>mżawka pojawia się</i>	<i>la bruine apparaît</i>
[Operacje akcesor :]	[Opérations accesseur :]
<i>mżawka ogranicza widoczność</i>	<i>la bruine réduit la visibilité</i>
<i>mżawka orzeźwia</i>	<i>la bruine rafraîchit</i>
<i>mżawka pada</i>	<i>la bruine tombe</i>
<i>mżawka przechodzi, zmienia się w deszcz</i>	<i>la bruine se change en pluie</i>
<i>mżawka ustaje</i>	<i>la bruine cesse</i>
<i>mżawka zwilża ziemię</i>	<i>la bruine humecte le terrain</i>
[Operacje manipulator :]	[Opérations manipulateurs :]
<i>przewidzieć mżawkę</i>	<i>prévoir la bruine</i>
<i>odbywać się w mżawce</i>	<i>se dérouler sous la bruine</i>
<i>iść w mżawce</i>	<i>marcher sous la bruine</i>
[Ekstensja :]	[Extensions :]
—	—

Voyons maintenant les rapports de l'approche orientée-objets avec d'autres conceptions portant sur la traduction automatique telles que : *WordNet, les classes d'objets, la notion de qualia structure et de fonction lexicale*.

Parmi une grande diversité de ressources lexicales WordNet constitue un exemple intéressant parce que c'est une base lexicale conçue pour le support électronique. C'est une base de données lexicographiques de l'anglais développée par le Cognitive Science Laboratory à Princeton University. Cette base lexicale contient à peu près 144 000 d'entrées constituées par des noms, des verbes, des adjectifs. Le principal objectif de WordNet est de présenter comment les sens des mots s'organisent les uns par rapport aux autres. Ce qui est le plus révélateur et original dans cette conception c'est la façon d'organiser les mots. WordNet constitue un réseau lexical où un sens se définit par les relations de proximité ou de contraste qu'il entretient avec les autres mots, ainsi est-il caractérisé par la place qu'il occupe dans tout le réseau. Dans le WordNet tous les hyponymes sont connectés à leurs types supérieurs alors WordNet constitue un véritable système d'héritage lexical où chaque mot hérite de sa hyperclasse les traits caractéristiques.

Tous les mots sont organisés en ensemble de synonymes appelés « synsets ». Les synsets sont liés les uns aux autres par différentes relations telles que : *hyperonyme homonyme, spécifique, générique, partie-tout*. Chaque synset fait partie d'un autre et ainsi de suite jusqu'au synset qui ne possède pas de classe supérieure, nommé « unique beginner ».

Voyons comment se présente la hiérarchisation du mot *bruine* (*drizzle*):

- drizzle** (very light rain; stronger than mist but less than a shower)
 - => rain, rainfall (water falling in drops from vapor condensed in the atmosphere)
 - => precipitation, downfall (the falling to earth of any form of water (rain or snow or hail or sleet or mist))
 - => weather, weather condition, atmospheric condition (the meteorological conditions: temperature and wind and clouds and precipitation)
 - => atmospheric phenomenon (a physical phenomenon associated with the atmosphere)
 - => physical phenomenon
 - => natural phenomenon (all non artificial phenomena)

Les auteurs du WordNet ont sélectionné 25 catégories de «unique beginners» telles que: *animal*, *artefact*, *corps*, *nourriture*, *plante*, *substance*, *temps*, *phénomène naturel*.

L'approche orientée-objets s'appuie sur cette division conceptuelle proposée dans WordNet. Les 25 catégories de noms spécifiées par WordNet constituent nos superclasses, dont nous ferons le point de départ de l'analyse de notre corpus.

Les relations de hiérarchisation qui vont être spécifiées dans l'approche orientée-objets sont similaires à la hiérarchisation faite dans WordNet :

- Puisque la bruine est une sorte de pluie*
- Puisque la pluie est une sorte de précipitation*
- Puisque la précipitation est une sorte de phénomène météorologique*
- Puisque le phénomène météorologique est une sorte de phénomène atmosphérique*
- Puisque le phénomène atmosphérique est une sorte de phénomène naturel*

Ainsi :

- (f) **bruine**
 - ⇒ pluie
 - ⇒ précipitation
 - ⇒ phénomène météorologique
 - ⇒ phénomène atmosphérique
 - ⇒ phénomène naturel

La caractérisation des entrées par l'intermédiaire des opérations et des attributs rend possible l'établissement et l'indication des classes d'objets pour

chaque nom. La description spécifie les sous-classes et les superclasses pour chaque nom en question :

bruine	
<i>superclasse 1</i>	pluie
<i>superclasse 2</i>	précipitation
<i>superclasse 3</i>	phénomène météorologique
<i>superclasse 4</i>	phénomène atmosphérique
<i>superclasse 5</i>	phénomène naturel

Le module présenté s'apparente en quelques points à l'approche des classes d'objets à la manière de Gaston Gross. Rappelons que la notion de « classe d'objet » conçue à des fins de traduction automatique renvoie à des ensembles de mots établis avant tout selon leur homogénéité sémantique. Les classes d'objets équivalent à des ensembles qui ont en commun un certain nombre d'attributs et d'opérations. Chaque classe est définie par des adjectifs, des verbes ou des noms que l'on indique comme les prédictats appropriés pour la classe en question. Ainsi la classe de « phénomènes naturels » est-elle caractérisée par les prédictats tels que : *avoir lieu, il y a, survenir, se produire*.

La notion de classes d'objets montre comment cet outil rend compte de la polysémie des prédictats verbaux, mieux que les trait traditionnellement utilisés dans ce but *humain, concret, abstrait, animé, inanimé*.

Les deux conceptions convergent vers le même résultat : l'établissement des classes d'objets linguistiques, c'est-à-dire déterminées à travers la langue et non pas fondées sur la division conceptuelle du monde. Dans ces deux conceptions les classes de noms sont établies en fonction des opérateurs qui l'accompagnent dans des situations évoquées.

Le partage des opérations en prédictateurs-constructeur, prédictateurs-accesseur et prédictateurs-manipulateur correspond aux rôles qui ont été spécifiés par Pustejovsky dans son modèle de *qualia structure* – basé sur les modes d'explication d'Aristote.

La *qualia structure* est un des éléments de représentation sémantique des noms (nous en avons quatre – structure argumentale, structure de l'événement, structure *qualia* et structure de l'héritage lexical) et permet de décrire les sens des noms. Cette structure renvoie à quatre rôles fondamentaux :

1. Le *rôle constitutif* qui définit la relation entre un objet et ses constituants.
2. Le *rôle formel* qui permet de distinguer l'objet parmi d'autres objets (*orientation, taille, forme...*).
3. Le *rôle télique* qui spécifie la fonction et le but de l'objet.
4. Le *rôle agentif* qui précise ce qui est à l'origine de l'objet.

La structure des *qualia* permet de rapprocher les phénomènes de « sens en contexte ». Elle fournit non seulement le sens des mots mais aussi suggère de différentes interprétations de ce mot en contexte. Voyons un exemple :

chêne

S-QUA	=	objet_naturel
CONST	=	<i>partie_du végétal (x)</i> (<i>tronc, feuille, branche</i>)
FORMEL	=	<i>objet naturel</i>
TÉLIQUE	=	<i>se développer (x), s'affruster (x), verdire (x)</i> <i>se mettre à fruit (x)</i>
AGENT	=	<i>objet naturel</i> <i>arroser (x)</i>

Tous ces rôles qu'englobe *qualia* sont reflétés par la structuration de nos opérations.

Les relations *Partie-tout* qui fournissent les données entre l'objet et ses constituants sont de même type que les informations qui exprime le rôle *Constitutif* chez Pustejovsky.

Le rôle *Formel* qui relie les différents niveaux est rendu par les relations entre une classe d'objets et sa classe supérieure.

Les informations apportées par le rôle *Agentif* sont fournies par les opérations et les résultats du prédicateur-manipulateur.

Les informations que nous apporte le rôle *Télique* correspondent aux opérations fournies par le prédicateur-accesseur. Mais dans le cas du mot *chêne*, comme d'ailleurs dans d'autres cas où le mot n'est pas un artefact mais un objet naturel ou un événement, ce rôle reste difficile à préciser (est-il possible de le faire?). Le rôle télique a été conçu avant tout pour les artefacts, donc pour les objets physiques qui sont produits par l'homme. Ainsi, il est impossible de déterminer ce rôle dans d'autres cas. En effet nous ne sommes pas capables de présenter la façon dont l'objet a été créé et de préciser ce qui en est à l'origine.

Le schéma descriptif que l'on propose dans la description du type orientée-objets structure et organise des informations caractérisant une classe donnée, de cette façon il s'apparente à la conception du *frame*.

La notion du frame est née dans les années soixante-dix et trouve son origine dans la psychologie. C'est une théorie développée par Minsky qui présente la façon de percevoir le savoir. L'essentiel de cette théorie montre que les gens utilisent des structures de connaissances, des ensembles d'informations structurées pour comprendre de nouvelles situations et pour agir. Ces structures sont constituées à partir d'expériences déjà vécues. Ce cadre référentiel sélectionné dans la mémoire est appelé *frame*. Pour pouvoir s'adapter à un grand nombre de situations nouvelles, un frame est dynamique, il peut être changé et transformé en fonction des détails.

Dans la linguistique le module du frame a pour objectif de présenter toutes les configurations possibles des éléments avec tous ces partenaires conceptuels. Ces configurations sont établies sur base des liens conceptuels qui réunissent les objets apparentés du point de vue sémantique, p.ex.: *dictionnaire, livre, roman*. Dans l'approche orientée-objets ce type d'informations associant les objets apparentés sémantiquement est fourni par la partie *Synonyme*:

bruine, crachin
tempête, ouragan, orage

Dans l'approche présentée basée sur la notion des classes d'objets, nous pouvons retrouver encore un élément important: le concept de fonction lexicale de la théorie Sens-Texe de I. Mel'čuk.

Rappelons que la notion de la fonction lexicale a été créée pour faire face à toutes les expressions phraséologiques qu'on ne peut pas prévoir dans la langue et que la linguistique ne sait pas présenter de façon logique – appelées *collocations*. La nouveauté du dictionnaire proposé par I. Mel'čuk consiste en une technique particulière utilisée pour la description de ces collocations.

La fonction lexicale (FL) est une fonction qui associe à un mot *X* un autre mot ou un groupe des mots *Y* (appelés valeurs de la fonction) qui sont liés sémantiquement à *X*. I. Mel'čuk et A. Zolkovsky ont distingué environ 60 FL. Les FL servent à préciser le type de relation entre l'objet en question et l'opération. Parmi les FL, certaines acceptent un grand nombre d'arguments et possèdent un assez grand nombre de valeurs différentes comme les fonctions *Oper, Func, Labor*.

Le même type de spécification des relations nous retrouvons dans l'approche orientée-objets en examinant le partage des opérations en trois types de prédicteurs – prédicteurs-constructeur, prédicteurs-accesseur et prédicteurs-manipulateur.

Analysons quelques exemples de fonctions lexicales :

Func_0 , Func_1 , est un verbe qui ne possède pas de contenu sémantique déterminé, qui prend le nom de la situation comme son sujet grammatical (SG) et le nom du premier, deuxième, ... participants de la situation comme son complément d'objet principal au cas où la situation a des actants.

Func_0 (*bruine*) = *tomber*
 Func_0 (*vent*) = *souffler*
 Func_0 (*brouillard*) = *apparaître*

Il est possible d'ajouter à ces fonctions dites « standard » différents sens p.ex. *de cesser* en formant ainsi des FL complexes.

FinFunc₀ (*pluie*) = cesser

FinFunc₀ (*vent*) = se taire

FinFunc₀ (*brouillard*) = se dissiper

Les données qui sont fournies par la fonction lexicale *Func* correspondent aux opérations du prédicteur-accesseur.

Oper, est un verbe support qui prend le nom du premier, deuxième participant de la situation pour son sujet grammatical et le mot clé comme complément d'objet.

Oper₁ (*pluie*) = regarder, recevoir

Oper₁ (*vent*) = entendre

Oper₁ (*brouillard*) = faire (*il fait du brouillard*)

Les données que nous apporte cette fonction sont exprimées par les informations du prédicteur-manipulateur.

Labor_{ij} est un verbe qui n'a pas de contenu sémantique déterminé et qui associe le nom du participant *i* comme son sujet grammatical au nom du participant *j* en tant que son complément d'objet principal et au nom de la situation donnée comme complément d'objet indirect :

Labor_{ij} (*pluie*) = être mouillé, être trempé (*qqn est trempé de pluie*)

Labor_{ij} (*vent*) = être hâlé (*qqn est hâlé par le vent*)

Labor_{ij} (*brouillard*) = rouler dans (*qqn roule dans le brouillard*)

Les informations qu'apporte la fonction *Labor* sont exprimées par les opérations du prédicteur-manipulateur.

Magn est une fonction lexicale qui correspond à un modificateur intensificateur des noms, des verbes.

Magn (*pluie*) = grosse, forte, diluvienne, torrentielle

Magn (*vent*) = fort, violent

Magn (*brouillard*) = à couper au couteau

Les données qui peuvent être exprimées à l'aide de la fonction *Magn*, dans la description orientée-objets sont fournies par les attributs.

Conclusion

Dans le présent article nous nous sommes concentrés sur la présentation de l'approche orientée-objets – conception proposée par professeur W. Banyś. Nous avons exposé l'idée de la conception elle-même en nous appuyant sur les exemples appartenant à la classe de *phénomènes naturels*. Ensuite nous avons essayé de révéler ce qui a servi de tremplin à cette théorie, à savoir d'autres théories lexicographiques traitant le problème de la traduction automatique.

Références

- Apresjan J., 2000: *Semantyka leksykalna. Synonimiczne środki języka*. Wrocław, Zakład Narodowy im. Ossolińskich.
- Banyś W., 2001: «Bases de données lexicales électroniques – une approche orientée-objets». *Neophilologica*, 15.
- Banyś W., 1995: „Kognitywna leksykografia komputacyjna: Teoria Słownika Generatywnego Pustejowsky'ego i Boguraeva”. *Bulletin de la Société Polonaise de Linguistique*.
- Clas A., Gross G., 2003: “Analyzing LSP for Automatic Data Processing the Language of Football”. *Language Research*, 1–19.
- Benoit H., Nazarenko A., Salem A., 1997: *Les linguistiques de corpus*. Paris, Armand Colin.
- Buvet A., Foucou P., 2000: «Extraction automatique de noms de professions sur le Web». *Linguisticae Investigationes*, 24.
- Giry-Schneider J., 1994: «Sélection et sémantique: Problèmes et modèles». *Langages [Larousse]*, 115, 5–14.
- Le Pesant D., Mathieu-Colas M., 1998: «Introduction aux classes d'objets». *Langages [Larousse]*, 131, 6–33.
- Mathieu-Colas M., 1998: «Illustration d'une classe d'objets: les voies de communication». *Langages [Larousse]*, 131, 77–90.
- Mel'čuk I., 1997: *Cours de morphologie générale*. Vol. 1. CNRS Editions.
- Paykin K., 2002: «Événements, états et substances: un essai météorologique». *Cahiers CHRONOS* [New York], 183–199.
- Pustejovski J., 1998: *The Generative Lexicon*. Cambridge, Mass., London, MIT Press.